Beyond Inclusive Fitness? On A Simple And General Explanation For The Evolution Of Altruism

Alejandro Rosas§

Two methodologies are available for calculating the fitness of altruism, namely inclusive fitness and neighbor modulated fitness. Inspired by the second methodology, Fletcher and Doebeli (2009) propose a new approach to the evolution of altruism, where assortment plays a fundamental role. Weak and reciprocal altruism appear as genuine cases of altruism in this new approach. In this paper I argue that the approach implies a new concept of altruism: a cooperative behavior is altruistic whenever it requires positive assortment between altruists to evolve. Moreover, assortment between altruists is controlled by traits that evolve by natural selection in the individual altruists. The role of assortment, and its being controlled by individual traits that co-evolve with altruistic ones, is the fundamental new insight promoted by the neighbor modulated fitness approach. I also examine and reply to the criticisms of inclusive fitness theorists against this new perspective on altruism.

1. Concepts of Altruism

Altruism is a central concept in evolutionary biology. Evolutionary biologists still disagree about its meaning (E.O. Wilson 2005; Fletcher et al. 2006; D.S. Wilson 2008; Foster et al. 2006a, b; West et al. 2007a, 2008). Semantic disagreement appears to be quite robust and not easily overcome by attempts at clarification, suggesting that substantive conceptual issues lurk in the background. Briefly, group selection theorists define altruism as any trait that makes altruists losers to selfish traits within groups, and makes groups of altruists fitter than groups of non-altruists. Inclusive fitness theorists reject a definition based on within- and between-group fitness. Traits are altruistic only if they cause a direct and absolute fitness loss to the donor. The latter definition is more restrictive and rejects as cases of altruism behaviors that are accepted by the former. Fletcher and Doebeli (2009) recently proposed a simple, direct and individually based fitness approach, which they claim returns to first principles: carriers of the genotype of interest “must, on average, end up with more net direct fitness benefits than average population members.” This seductively simple proposal uses the concept of assortment to explain how diverse kinds of altruists end up on average with more net fitness than their non-altruistic rivals. In this paper I shall argue that their approach implies a new concept of altruism that contrasts with and improves on the concept of the inclusive fitness approach.
2. A New Concept

Explanations of the evolution of altruism have both a conceptual side and a mathematical “accounting” side for calculating the fitness of an altruistic trait. There is consensus that different accounting systems such as those practiced under the perspective of inclusive fitness (IF) or group selection or neighbor modulated fitness (NMF) are mathematically equivalent and that pluralism in accounting is acceptable as long as it leads to useful and tractable models (West et al. 2007a; Wilson 2008). However, disagreement over the conceptual side of the issue has not settled down. Group selection theorists have claimed that their way of partitioning fitness in within- and between-group components alone does justice to the deep causal structure of the evolutionary process. IF theorists disagree; many claim that any process that can be understood from the perspective of group or multilevel selection can be better captured within the IF approach (West et al. 2007a, 2008). Recently, a new contender in this conceptual disagreement has appeared: Fletcher and Doebeli (2009) developed a view of weak and reciprocal altruism (to be defined below) as genuine cases of altruism within the NMF approach. In their understanding of this approach, assortment between altruists replaces relatedness and achieves greater generality, justifying the inclusion of cases that are rejected by IF theorists.

This view of the NMF approach provokes the resistance of inclusive fitness theorists. They acknowledge that many evolutionary biologists use the NMF methodology for reasons of convenience, because it has often advantages for constructing the mathematical models. But since it is mathematically equivalent to the IF methodology, they reject the suggestion that it implies a new concept. For them, weak and reciprocal altruism are non-altruistic forms of cooperation (Foster et al. 2006a, b; West et al. 2007a, 2008). Their inclusion as instances of altruism replicates the same semantic confusions that have been advanced by group selection theorists.

2.1 Equivalent Methodologies

I begin by briefly explaining how the IF and NMF methodologies are mathematically equivalent, even though they give different interpretations of the fitness of altruists and how they evolve. Both methodologies go back to Hamilton (1964). They differ on how they partition and calculate the fitness of an altruistic behavior. Though his famous rule is mathematically appropriate in both approaches,

\[ rb - c > 0 \]

the symbols ‘r’ and ‘b’ receive different readings. In the IF methodology the focal individual is interpreted as a donor. Its action produces a negative effect on its fitness, –c, sometimes called the direct fitness effect, and an indirect benefit b, which is allocated to the donor when the theorist takes the gene’s-eye view and realizes that the donor reproduces its genes through recipients that are genetically related, either by descent or otherwise. The benefit b conferred on recipients, multiplied by the coefficient of relatedness r, gives the indirect benefit to the donor. In contrast, the focal individual in the NMF approach is both recipient and donor, and the approach focuses on its “personal” fitness. It only counts its own direct offspring, which result from the sum of the direct effect of its altruism on itself, –c, and the indirect benefit (rb) received from altruistic neighbors: ‘b’ measures how much benefit flows from altruists, ‘r’ measures how much of it falls on the focal altruist rather than on non-altruists and is read as the coefficient of assortment between the altruistic gene of the focal individual and helping behaviors in its average interaction environment. Sometimes, but not always, assortment will imply interaction with genetic relatives. Figure 1 illustrates the difference between both approaches.
2.2 Assortment: The Key Issue

There is consensus that the two accounting systems are mathematically equivalent when they are used to calculate the fitness of altruism between relatives (West et al. 2007a; Gardner and Foster 2008). But an obvious problem for the inclusive fitness approach is that it cannot explain the evolution of altruism when donors and recipients are not genetically related, either by descent or otherwise. On the IF approach, the fitness of the altruist contains an indirect component that depends on benefits conferred on genetically related recipients. On the NMF approach, the fitness of the altruist depends on receiving the benefits of altruism, independently of their genetic origin. All that matters here is the assortment between the altruist’s gene or genes and help received.

According to Fletcher and Zwick (2006) and to Fletcher and Doebeli (2009) the most general explanation for the evolution of altruism is positive assortment between the altruistic gene of the focal individual and help received. Assortment is positive whenever the probability that an altruist receives an altruistic benefit is higher than the probability that a non-altruist receives the same benefit. This view achieves greater generality in two ways. First, altruism is explained as long as altruists receive altruistic donations, even if these do not come from genetically related individuals. The donations depend on assortment, which is a more general phenomenon than relatedness and includes it as a special case. Assortment can promote altruism in absence of relatedness (Fletcher and Zwick 2006). Second, it provides a common explanation for both strong (\(b < c\)) and weak (\(b > c\)) altruism (where \(b\) and \(c\) represent the benefit and cost for the focal altruist of its altruistic action), such that the reason for denying the later as genuine altruism disappears. In both cases the trait evolves when the probability that the altruist receives an altruistic benefit is higher than the probability that a non-altruist receives the same benefit. In the special case of weak altruism, this probability is higher even when individuals are distributed randomly into groups with respect to genotype.

Inclusive fitness theorists have been very critical towards these consequences. They deny that behaviors like weak and reciprocal altruism, which can exist among non-relatives, should count as true altruism. Usually they criticize group selection theorists for endorsing a concept of altruism that leads to these claims, thus confounding, in their view, altruism with cooperation (Foster et al. 2006b; Lehmann and Keller 2006; West et al. 2007a). However, their criticism disregards the fundamental role of assortment in the evolution of altruism, which is the main conceptual contribution of the NMF approach. An argument based on
assortment and formulated independently of the group selection perspective justifies the inclusion of weak and reciprocal altruism as genuine cases. I explain this argument below, first in relation to weak and then in relation to reciprocal altruism.

3. Weak Altruism

The weak altruist donates at a cost of \( c \) a benefit \( b \) to every group member including itself, such that \( b > c \). Based on the lower relative fitness of weak altruists within groups, Wilson (1979) argued that it requires an explanation in terms of group selection. But weak altruism can evolve in a subdivided population even when individuals are distributed randomly into groups with respect to genotype. In this case, the cooperative environment of altruists and non-altruists is exactly the same; and yet weak altruists end up, on average, with higher fitness than the non-altruists. This is taken to mean that weak altruism is not true altruism (Nunney 1985). West et al. (2007a) denounce the label “weak altruism” as a source of confusion, because weak altruists are fitter than non-altruists in the population as a whole when random interaction is assumed. This is primarily meant as an argument against how group selection theorists view weak altruism. It cannot be used against the NMF approach, however, because NMF measures the fitness of altruists in relation to the population as a whole, and it is always higher than the fitness of non-altruists for all cases of evolved altruism. On the NMF view, the fundamental fact is positive assortment, i.e., that the probability of receiving help is higher for an altruist than it is for a non-altruist. This suggests a redefinition of altruism: a trait is altruistic if it needs positive assortment to evolve. This is a sufficient condition and probably also necessary: cooperative traits are non-altruistic if they evolve even by full negative assortment. This includes all those traits that donate benefits as by-products of selfish effects and where the benefit to the actor is greater than the benefit to recipients (Sachs et al. 2004). Sentinel behavior in meerkats, *Suricata suricatta*, is an example according to some interpretations (Clutton-Brock et al. 1999).

Weak altruism fulfills the new definition when positive assortment is understood as above (Section 2.2). Positive assortment is realized in weak altruism by random genotype distribution because the focal altruist donates benefits also to self. In Fletcher and Doebeli’s (2009) notation:

\[
\frac{e_C b + b N - c}{N} > \frac{e_D b}{N}
\]

where \( e_C \) is the number of cooperators in the average interaction group of a focal altruist, \( e_D \) is the number of cooperators in the average interaction group of a focal non-altruist, \( b \) is the total benefit conferred on the group, \( N \) is the number of players in a group (the same for all groups). Note that this equation is best interpreted within the NMF approach: the LHS represents the sum of cost and benefit to the personal fitness of the focal altruist, the RHS represents the benefit to the non-altruist. This transforms into the following equation in the format of Hamilton’s rule:

\[
\left( \frac{e_C + 1}{N} - \frac{e_D}{N} \right) b - c > 0
\]

where \( \left( \frac{e_C + 1}{N} - \frac{e_D}{N} \right) \) is equivalent to \( r \), the coefficient of relatedness.

Now suppose that individuals are distributed randomly into groups with respect to genotype, which means that \( e_C = e_D \). In spite of this fact, the coefficient of relatedness or assortment is positive: \( \frac{1}{N} \). This insight about positive assortment was expressed by Pepper (2000) in relation to average relatedness of recipients to actors for “whole group” as opposed to “other only” altruistic traits in infinite populations.
subdivided randomly into groups with respect to genotype. Here relatedness is positive and equal to \( \frac{1}{n} \), where \( n \) is the number of individuals in a group. It is significant if groups are small. If the trait in question confers benefits on others but not on self (“other-only”), \( r \) is positive only when \( e_C > e_D \). In both cases, however, positive assortment between the carriers of the altruistic genotype and the amount of helping in their environment explains the evolution of altruism.

3.1 An Objection

Adopting the NMF approach suggests that there is only one fundamental fact in the evolution of altruism: that the probability of receiving help is higher for a focal altruist than for a focal non-altruist. From this perspective, weak altruism is as genuine as strong altruism and there is no reason to believe that true altruism exists only towards genetically related individuals.

Against this view, some evolutionary biologists argue that weak altruism evolves in virtue of the benefit that the weak altruist confers on self. Since this benefit is due to the altruist’s own behavior, it is a direct fitness effect and different from a behavior that evolves in virtue of an indirect benefit obtained through related individuals. This fundamental fact is revealed when applying the “mutation test” (Nunney 1985). If you mutate a non-altruist into a weak altruist in a population subdivided randomly into groups, the weak altruist will increase its fitness relative to the whole population and altruism will evolve.

This is true, but it assumes that random mixing will remain a static feature of the environment no matter what. The crucial question is what will stably maintain assortment between altruists at the required level to favor the evolution of altruism. An individual that mutates into a weak altruist to benefit itself may motivate one or more former altruists to mutate into non-altruists to exploit the newly developed altruist. This may alter the minimum condition for positive assortment of weakly altruistic traits, \( e_C = e_D \). Since altruists of any sort create a selection pressure in others to exploit them, evolutionary thinking cannot assume that the minimum condition for a positive coefficient of assortment of weakly altruistic traits will remain constant. Special traits in individuals may be required to maintain the minimum condition. It is crucial, moreover, that these special traits evolve by natural selection in the altruists themselves to make them evolutionary stable. The sections that follow justify this claim. Attention to the role of traits controlling assortment in individual altruists will also allow us to neutralize the criticisms that IF theorists have expressed against the NMF approach as developed in Fletcher and Doebeli (2009).

3.2 Assortment Co-evolves with Altruistic Traits

It is necessary to consider the different biological mechanisms that are capable of creating or maintaining positive assortment when thinking about the evolution of altruism (Fletcher and Doebeli 2009: 13). But, more importantly, the mechanisms that produce or control assortment must reliably emerge as traits of the individual altruist, if altruism is to evolve. Assortment and population structure are usually facilitated or influenced by traits in the individual organisms. It may happen through cognitive abilities, i.e., when individuals are selective or choosy in their interactions. But non-cognitive traits can facilitate spatial distributions that capitalize on genealogical relatedness to create assortment. Limited dispersal, for example, is plausibly facilitated by individual traits. The concept of control seems legitimate if we have reason to believe that traits facilitating assortment evolved to do precisely this, because they allow altruists to reap the benefits of cooperation. When the benefit conferred on neighbors is greater than the cost of production, mutual cooperation represents a win-win situation. It will be favored by natural selection, provided that the problem of assortment can be solved. If a trait evolves to fixation because it facilitates assortment between altruists, we can say that it controls assortment.

Sometimes it is very difficult to resist the claim that a trait evolved to facilitate or control assortment. This is the case when dispersal adopts sophisticated forms. While limited dispersal increases the kinship of social partners and favors altruism, it also intensifies local competition among the extra progeny generated by altruism, who compete locally for resources in altruistic groups. In order to avoid intense competition...
between relatives, progeny is dispersed in buds, i.e., in small groups in some species (Wilson and Pollock 1992; Goodnight 1992; Gardner and West 2006). In this way the negative effects of limited dispersal are avoided while the positive ones are maintained. The long-range dispersal of buds reduces local competition among the extra progeny generated by altruism, yet kinship is maintained as progeny disperses together in groups. It seems safe to assume that this dispersal pattern is the work of natural selection. Relatedness by itself cannot produce assortment unless is goes together with whatever pattern of dispersal is required. Traits that control assortment co-evolve with altruism and are part of the phenomenon.

4. Beyond the Paradox

Besides implying different ways of describing and explaining the process supporting the evolution of altruism, adopting the NMF approach takes us to realize a further fact, not sufficiently highlighted by Fletcher and Doebeli (2009). This is the fact that traits evolve in the altruist to control assortment. This is the fundamental conceptual contribution of the NMF approach and its deepest contrast to IF. The IF methodology goes together with the idea that there is a paradox in the evolution of altruism: altruists really lose fitness as individuals, and the genes they carry only evolve because they reproduce through relatives. It is tempting to believe that theorists that defend the superiority of inclusive fitness over group selection – and yet resist the conceptual conclusions drawn from the NMF methodology – are under the grip of an illusion, strongly suggested in passages by Hamilton like this one: “…a gene may receive positive selection even though disadvantageous to its bearers if it causes them to confer sufficiently large advantages on relatives” (Hamilton 1964, 17). Fletcher and Doebeli (2009) get close to proposing a conceptual revolution when they describe this illusion, and its remedy, in the following way:

This quote implies that there are two types of individuals who experience two very different fitness effects: bearers, who suffer disadvantages, and relatives, who garner advantages. But of course, the only relatives that matter when viewing this situation from the inclusive fitness perspective are those that are themselves bearers (of the altruistic gene) (18).

The claim here is that the inclusive fitness approach nourishes the belief in a fundamental asymmetry. Altruists are losers in virtue of their altruism. They suffer a net fitness loss, while their altruistic gene spreads through their relatives’ augmented fitness. The invitation is to abandon this asymmetry and to embrace the fundamental symmetry of assortment. If you think that the altruistic gene will spread in spite of the fact that donors are losers, you are victim of an illusion: although some donors will be losers, they are not losers on average, and cannot be, if altruism evolves. Through assortment, donors are recipients and recipients are donors. The altruistic gene spreads only if, and only because, donors are compensated for their costs through assortment with other altruists. Simply put, the altruistic gene will never evolve unless it spreads through the personal fitness of altruistic donors. By controlling assortment, they increase their personal fitness, even as altruists, without paradox. The invitation, therefore, is to see the personal or direct fitness approach suggested by the NMF methodology as the fundamental and more general perspective. Donors are selected to receive, on average, more fitness benefits than what they pay out (Fletcher and Zwick 2006, 254).

It is important to realize that altruistic donors are winners on average. An illusion of asymmetry may arise because the actual distribution of costs and benefits can be highly asymmetric among donors. In the extreme cases of suicidal altruism and sterility in eusocial insects, some donors will receive notoriously less than other donors in terms of personal fitness. The expression of altruism in such extreme cases is probabilistic, for at least some donors must enjoy personal fitness benefits. And on average, carriers of the altruistic genotype must win over non-carriers.

5. Altruists Control Assortment

It is useful to note a gradual conceptual transition between altruism through assortment with kin on one hand, and reciprocal altruism on the other. Besides patterns of dispersal, kin recognition is another mechanism for assorting genotypes. Assortment is thus independent of spatial proximity and the expression
of altruism becomes conditional on the positive detection of kin: the behavior will be expressed whenever the detection mechanism goes to the positive state. That is, the detection mechanism must be hard-wired in the appropriate way to the expression of the behavior. A similar conditionality applies to green-beards, a mechanism that recognizes the specific allele for helping independently of whole-genome (kin) relatedness. Both mechanisms are vulnerable and their role in the evolution of altruism is currently debated. However, ways have been suggested to circumvent the problems (Gardner and West 2007). When the marker for recognition is the cooperative behavior itself, the mechanism develops into reciprocal altruism, which makes expression of altruism conditional on the perception of altruistic responses by others (Axelrod and Hamilton 1981). Reciprocal altruism differs from the green-beard mechanism because the genotype of the helpers is irrelevant in the former. All that matters is their helping phenotype.

The gradual transition from kin altruism to reciprocal altruism suggests a fundamental similarity between them. Yet an argument based on the mutation test questions the altruistic character of reciprocal altruism (Andy Gardner, personal communication). The argument says: Mutate an original defector (OD) into an altruist; in the presence of a reciprocator, the OD receives a fitness benefit not otherwise received. It owes this to its new behavior and therefore it confers a benefit on itself: this is cooperation and not altruism. Contrast with the mutation of an OD to an altruist among kin: the new altruist pays a cost, and receives no benefit from relatives that it would not have received before mutating. True altruists receive benefits from kin that are not due to their own actions. Thus, these are truly indirect effects, and altruism is genuine.

But this argument does not go deep enough. The described difference emerges only if you depict the environment of the OD mutated into a reciprocator as already an environment of reciprocators, and the environment of an OD mutated into a kin altruist as already an environment of kin. This happens by fiat in the argument, but why should this be so? This is only so if the altruistic trait has co-evolved with traits that control assortment and is responsive and conditioned to assortment. The mutated reciprocal altruist will gain benefits due to its own behavior only if it can control its vicinity to reciprocators. Reciprocal altruists need more besides conditional behavior to succeed. If assortment through conditional play fails consistently because neighbors are mobile cheaters (Enquist and Leimar 1993) or punishers (Clutton-Brock and Parker 1995), a mutant reciprocal altruist in a population of social individuals will not gain any benefits and will not spread; not for lack of conditionality, but for lack of assortment with like.

Similarly, the mutated kin altruist would have enjoyed the donations of other altruists before mutating if it had already assorted with them. But assortment requires the expression of a trait different from altruism (be it a pattern of dispersal, be it recognition and its marker). The altruist controls assortment through this trait and thus controls the benefits received in a way entirely comparable to the reciprocal altruist. At the more fundamental level, altruists that have stably evolved will have done so because they control assortment; and in this sense the benefits they receive are always due to their own traits or action. Altruists of any sort create a selection pressure in others to exploit them and profit at their expense. This is why the natural selection of altruism must also be the natural selection for traits that control positive assortment between altruists and create the required population structure. At a fundamental level, assortment is always controlled by the altruists themselves, and the benefits received by the altruist are due to their own traits. This is the fundamental conceptual insight underlying the NMF perspective.

### 6. Altruists Maximize Personal Fitness

The issue of control surfaces in an argument by which some evolutionary theorists attempt to spell out what is special about the inclusive fitness approach. They believe, namely, that if you stick to the Darwinian idea that organisms are designed to maximize something, this something is inclusive fitness (Grafen 2006; West et al. 2007a; Gardner and Foster 2008), and not neighbor modulated fitness, simply because organisms control the former but not the latter. Inclusive fitness is superior to neighbor modulated fitness, not mathematically, but conceptually. It is the authentic Hamiltonian achievement and the proper Darwinian way of understanding social organisms.
The claim is that organisms are not designed to maximize their personal fitness in social exchanges, because they have no control or command over effects due to the helping behavior of others (Gardner and Foster 2008, 7-8). In contrast, they do control their own helping, which includes the direct fitness effect on the helper and the indirect effect achieved by helping relatives. Both effects add to inclusive fitness. This view is inaccurate as can already be inferred from the preceding discussion. First, the indirect fitness effects – and with them inclusive fitness – are not controlled by the mere act of helping. How often help will fall on relatives (inclusive fitness) depends on mechanisms that control assortment with kin. These mechanisms are not included in the helping trait, but are rather separate mechanisms like dispersal patterns and kin recognition. Second, if controlling assortment is the clue to controlling inclusive fitness and if the organism can be credited with it, the organism controls inclusive fitness and neighbor-modulated fitness in one move. The fundamental symmetry of assortment guarantees that the helper is also helped. From this perspective, the altruist controls its personal fitness, because the trait that controls assortment controls the effects that increase the personal fitness of the helper.

7. Conclusion and Summary

Inspired by the NMF methodology to calculate the fitness of altruistic traits, Fletcher and Zwick (2006) and Fletcher and Doebeli (2009) argue that the NMF approach is superior to the IF approach because it explains in a more general way the evolution of altruism, including cases where it evolves between non-genetically related individuals. This approach promotes a new concept of altruism where assortment plays a fundamental role. Weak and reciprocal altruism are genuine cases according to this concept. The approach also leads to the insight that traits for controlling assortment co-evolve with traits for altruism. The role of assortment and its control through traits in the individual altruists are fundamental facts outside the focus of the IF approach. Rather, the concept of altruism implied in the IF approach is an illusion that disregards the fundamental symmetry of assortment. The inclusive fitness approach mistakenly nourishes the idea that the altruistic organism is designed to control inclusive fitness but not neighbor modulated fitness. But the role of assortment and its symmetrical nature implies that it is designed to control both in one move. Whenever altruism evolves, a trait for controlling assortment co-evolves in altruists. This is suitably grasped under the NMF approach.

Literature cited


Acknowledgments

I would like to thank the Konrad Lorenz Institute for Evolution and Cognition Research and the Universidad Nacional de Colombia for financial support. Thanks to Andy Gardner and Jeff Fletcher for critical comments on earlier drafts and to Jonathan Kaplan for helpful comments at the final stages.

Copyright © 2010 Author(s).
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ISSN 1949-0739